Characterization of a novel *Streptococcus thermophilus* rolling-circle plasmid used for vector construction

Abstract The complete nucleotide sequence of pER371, a native plasmid in *Streptococcus thermophilus* ST137, was determined. A putative open reading frame coding for a replication protein, Rep371, was identified. A characteristic promoter sequence and ribosome-binding site were found upstream of *rep371*. Rep371 (247 amino acid residues) does not show homology with RepA and RepS of the small *S. thermophilus* cryptic plasmids pST1-No.29 and pST1 respectively. The plus-origin sequence and Rep371 are highly homologous to the corresponding elements of the *Staphylococcus aureus* plasmids pC194 and pSK89. A novel 140-nucleotide palindromic minus-origin sequence, which is structurally similar but does not show sequence homology to the *palA* region of pC194, was identified in pER371. A palindromic sequence capable of forming a putative hairpin structure was identified and subsequently recognized as being highly conserved among several lactococcal rolling-circle plasmids. Cloning vectors derived from pER371 should provide valuable gene-delivery vehicles for the genetic engineering of lactic acid bacteria.

Introduction

Streptococcus thermophilus is a lactic acid bacterium important to the dairy industry. It is used in the starter cultures for the production of yogurt and certain cheese varieties. As with the other microorganisms used in the industrial processes, research efforts have been intensive in developing new and/or improved strains through genetic engineering. Heterologous gene expression represents an avenue by which the target organism could be endowed with new and/or improved biological properties. In this respect, the desired genes are carried into and maintained in the host bacterium by a cloning or expression vector. The expression of heterologous genes in *S. thermophilus* has been reported from experiments using vectors derived from other bacteria (Constable and Mollet 1994; Jacobs et al. 1995; Mercenier 1990; Moineau et al. 1995; Mollet et al. 1993; Somkuti et al. 1991, 1993; Vaughan et al. 1996). The construction of homologous cloning vectors for *S. thermophilus* has been initiated recently in the interest of developing food-grade gene-delivery systems (Solaiman and Somkuti 1993; Somkuti et al. 1995). Since these vectors were built from the native plasmids found in various strains of *S. thermophilus*, systematic analyses of these plasmids should provide important information for the rational design and construction of the expression vectors for this industrial bacterium.

Native plasmids of lactic acid bacteria have been found to replicate via one of the following two mechanisms, i.e., theta-type and rolling-circle (Gruss and Ehrlich 1989; Novick 1989; Seegers et al. 1994). The theta-type plasmids are generally larger and more stably maintained in the host cells than are the rolling-circle DNA (Kiewiet et al. 1993; Krabbe et al. 1997; Renault et al. 1996). The occurrence of native plasmids in *S. thermophilus* has been documented (Herman and McKay 1985; Somkuti and Steinberg 1986a), but sequence determination and analysis of these plasmids are still limited (Hashiba et al. 1993; Janzen et al. 1992). In this communication, we report the complete sequence of a 2.67-kb *S. thermophilus* plasmid, pER371, and the analysis of its structural features.

D. K. Y. Solaiman (✉) · G. A. Somkuti
U.S. Department of Agriculture, ARS,
Eastern Regional Research Center,
600 East Mermaid Lane,
Wyndmoor, PA 19038, USA
e-mail: dsolaiman@arserrc.gov
Tel.: + 1-215-233-6476
Fax: + 1-215-233-6589

Materials and methods

Bacteria, growth conditions, plasmids

Escherichia coli DH5α was purchased from BRL Life Technologies Inc. (Gaithersburg, Md.). *S. thermophilus* ST137 (NRRL B-18818) is from our laboratory collection. *E. coli* was grown in Luria medium (1% w/v tryptone, 0.5% yeast extract, and 0.5% NaCl) at 37 °C, and *S. thermophilus* was cultured in a tryptone/yeast extract/lactose broth (Somkuti and Steinberg 1986b) at 37–42 °C. Selection pressure was maintained in *E. coli* transformants by adding ampicillin (100 μg/ml) to the growth medium. Plasmid pUC19 was obtained from BRL Life Technologies. pER371 was isolated from *S. thermophilus* ST137. This strain harbors two plasmids, pER371 (2.7 kb) and pER372 (14.75 kb) (Somkuti and Steinberg 1986a).

Molecular biology procedures

Restriction endonucleases, T4 DNA polymerase, *E. coli* DNA polymerase Klenow fragment, calf intestinal alkaline phosphatase, and T4 DNA ligase were purchased from the following commercial sources and used according to the protocols supplied: BRL Life Technologies, New England Biolabs (Beverly, Mass.), and United States Biochemical (Cleveland, Ohio). Transformation of competent *E. coli* DH5α was carried out by a heat-shock procedure described in the supplier's technical data sheet. Recombinant plasmids were isolated from *Escherichia coli* by an alkaline lysis method (Birnboim and Doly 1979). When needed, plasmids were further purified by using Qiagen Midi or Maxi kits (Qiagen Inc., Chatsworth, Calif.) or a *PERFECT*prep DNA preparation kit (5′ → 3′ Inc., Boulder, Colo.). Native plasmids of *S. thermophilus* ST137 were isolated by a previously described method (Somkuti and Steinberg 1986b). The smaller pER371 was separated from pER372 on an agarose gel by electrophoresis and subsequently eluted from the gel matrix by electroelution. Agarose gel electrophoresis was performed in TBE buffer (89 mM TRIS base, 89 mM boric acid, 2 mM NaEDTA). Electroelution was carried out in a SixPac GE200 Eluter (Hoefer Scientific Instrument, San Francisco, Calif) or an S&S Elutrap System (Schleicher & Schüll Inc., Keene, N.H.). Samples of the electroeluted pER371 were purified and concentrated with Elutip-d columns (Schleicher & Schüll Inc.).

DNA sequence analysis

The dideoxynucleotide chain-termination method (Sanger et al. 1977) was used to determine DNA sequences. Double-stranded plasmid DNA templates were purified by CsCl gradient ultracentrifugation (Ausubel et al. 1987) or by using the *PERFECT*prep DNA-binding matrix according to the protocol supplied (5′ → 3′ Inc.). The AutoRead and AutoCycle sequencing kits (Pharmacia Biotech, Piscataway, N.J.) were used to perform the sequencing reactions. Sequence data were collected and processed in an A.L.F. DNA sequencer and fragment analysis system (Pharmacia Biotech). Database searches were performed with the BLAST program using the default parameter settings (Altschul et al. 1990). Multiple sequence alignments were carried out using DNASIS Windows 2.1 (Hitachi Software Engineering America, San Bruno, Calif.) and the CLUSTAL-W (Thompson et al. 1994) programs. Potential hairpin structures were predicted using the DNASIS Windows 2.1 package.

Results

Sequence determination of pER371

In preparation for sequence determination, pER371 was linearized or fragmented and separately subcloned into an *E. coli* vector, pUC19, to produce the pUER-series plasmids. The restriction sites chosen for the subclonings were such that the contiguous sequences subsequently generated would have extensive overlapping regions. With these plasmids as templates, the sequencing reactions were then performed using M13 universal and reverse primers. The sequencing runs typically yield 200- to 500-base stretches, with 300-nucleotide contiguous sequences as the norms. The complete pER371 sequence was assembled from these contiguous sequences on the basis of the forward and reverse sequences with at least 100 bp overlap. The result shows that pER371 contains 2672 base pairs. The calculated G + C content of pER371 is 38.16 mol%. A restriction map of pER371, constructed on the basis of the complete sequence, is shown in Fig. 1A. Aside from the restriction sites that are too close to be identified by agarose gel electrophoresis analysis, the map of the sequenced pER371 agrees well with the previously observed restriction patterns of the plasmid. The pER371 sequence has been deposited in GenBank (accession number AF022180).
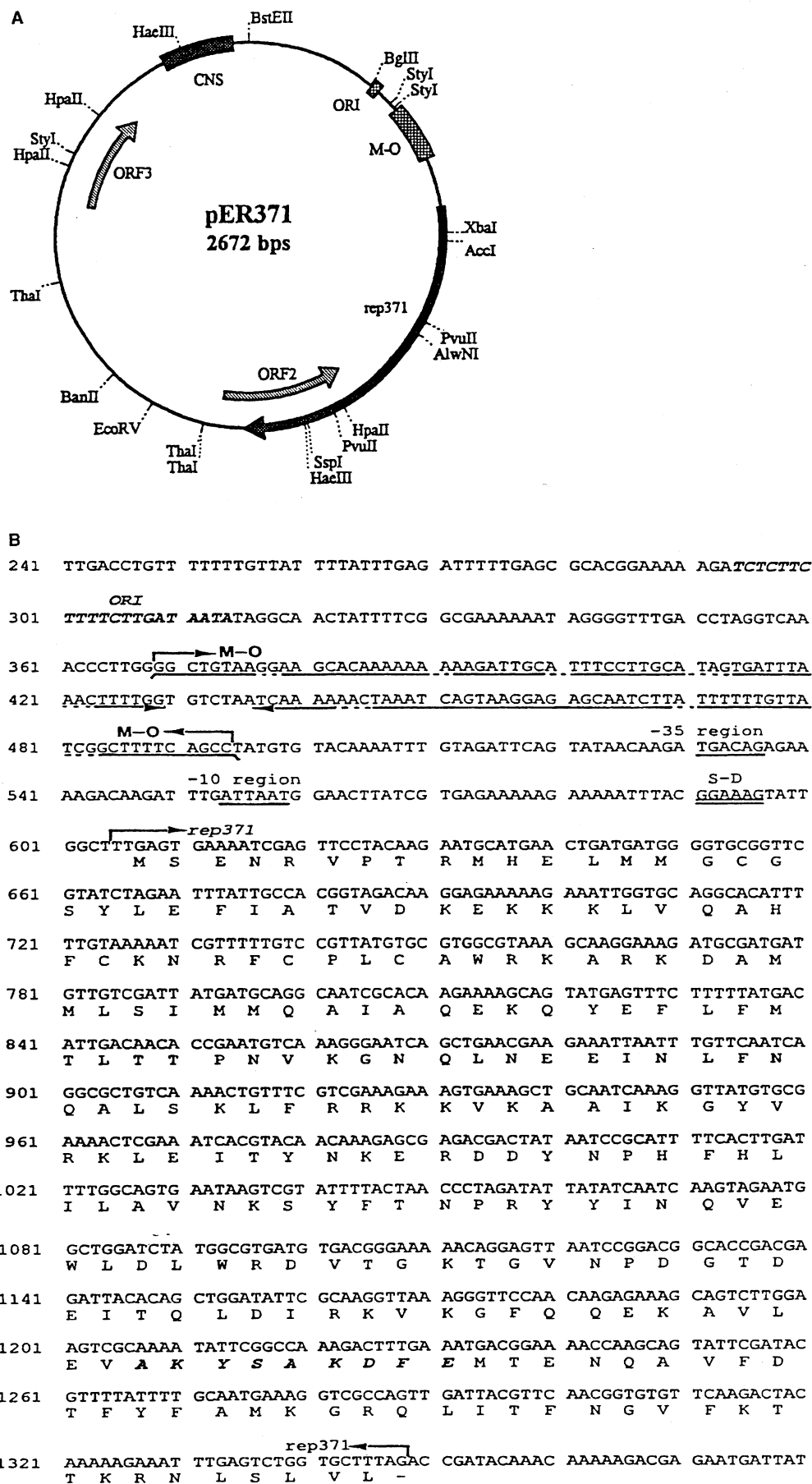
Characterization of the pER371 replication region

Computer analysis revealed the presence of a 744-bp open reading frame (ORF), designated as *rep371*, encompassing nucleotide numbers 605–1348. Figure 1B presents the sequence of *rep371* along with its 5′- and 3′- flanking regions. This ORF codes for a putative protein of 247 amino acid residues, with a calculated molecular mass of 29 018 Da and an isoelectric point of 9.85. A putative promoter region, constituted of the −10 (ATTAAT) and −35 (TGACAG) consensus sequences separated by a 17-nucleotide spacing, was identified upstream from *rep371*. A putative Shine-Dalgarno sequence (GGAAAG), similar to that reported earlier for the *repA* gene of *S. thermophilus* pST1 (Janzen et al. 1992), was located 8 bp upstream from the start codon of the ORF.

Basic local alignment search tool (BLAST) analysis (Altschul et al. 1990) of Rep371 showed that the putative polypeptide is homologous to the sequences of replication proteins associated with small plasmids of Gram-positive bacteria. The two sequences with the highest homologies to Rep371 are the Rep proteins of *Staphylococcus aureus* plasmids pC194 (Horinouchi and Weisblum 1982) and pSK89 (Littlejohn et al. 1991). Figure 2A shows the results of a multiple sequence-alignment analysis of the three replication proteins using the CLUSTAL-W program (Thompson et al. 1994). Nearly 66% of the 247 residues of Rep371 either are identical to or represent conservative substitution of the amino acid sequences of the pC194 and pSK89 replication proteins. The conserved active-site sequence associated with the pC194-class Rep proteins (Gruss and Ehrlich 1989) could be located in Rep371 between amino acid residues 201 and 209.

Examination of the DNA sequence in the region upstream from *rep371* revealed the presence of a se-

Fig. 1A, B Plasmid map and the sequence of the replication region of pER371. The first guanine in the unique *Bst*EII site (GGTGAC) is designated as nucleotide number 1. **A** The plasmid map summarizes sequence features identified in this study. **B** The −10 and −35 consensus promoter sequences and the putative Shine-Dalgarno signal of the Rep371 coding region (ORF1) are underlined and double-underlined respectively. The nick-site sequence of the plus-strand origin (ORI) and the conserved active-site amino acid sequence of Rep are both shown in italicized and bold face fonts. The underline arrows indicate the palindromic minus-strand origin (M-O) sequence of the plasmid; the intervening dotted lines are the non-complementary regions of the palindrome

A

B

```
241   TTGACCTGTT  TTTTTGTTAT  TTTATTTGAG  ATTTTTGAGC  GCACGGAAAA  AGATCTCTTC

      ORI
301   TTTTCTTGAT  AATATAGGCA  ACTATTTTCG  GCGAAAAAAT  AGGGGTTTGA  CCTAGGTCAA

                     M-O
361   ACCCTTGGGG  CTGTAAGGAA  GCACAAAAAA  AAAGATTGCA  TTTCCTTGCA  TAGTGATTTA

421   AACTTTTGGT  GTCTAATCAA  AAAACTAAAT  CAGTAAGGAG  AGCAATCTTA  TTTTTTGTTA

      M-O                                                       -35 region
481   TCGGCTTTTC  AGCCTATGTG  TACAAAATTT  GTAGATTCAG  TATAACAAGA  TGACAGAGAA

               -10 region                                       S-D
541   AAGACAAGAT  TTGATTAATG  GAACTTATCG  TGAGAAAAAG  AAAAATTTAC  GGAAAGTATT

                      rep371
601   GGCTTTGAGT  GAAAATCGAG  TTCCTACAAG  AATGCATGAA  CTGATGATGG  GGTGCGGTTC
          M   S    E   N   R    V   P   T    R   M   H   E    L   M   M    G   C   G

661   GTATCTAGAA  TTTATTGCCA  CGGTAGACAA  GGAGAAAAAG  AAATTGGTGC  AGGCACATTT
       S   Y   L   E    F   I   A    T   V   D    K   E   K   K    K   L   V    Q   A   H

721   TTGTAAAAAT  CGTTTTTGTC  CGTTATGTGC  GTGGCGTAAA  GCAAGGAAAG  ATGCGATGAT
       F   C   K   N    R   F   C    P   L   C    A   W   R   K    A   R   K    D   A   M

781   GTTGTCGATT  ATGATGCAGG  CAATCGCACA  AGAAAAGCAG  TATGAGTTTC  TTTTTATGAC
       M   L   S   I    M   M   Q    A   I   A    Q   E   K   Q    Y   E   F    L   F   M

841   ATTGACAACA  CCGAATGTCA  AAGGGAATCA  GCTGAACGAA  GAAATTAATT  TGTTCAATCA
       T   L   T   T    P   N   V    K   G   N    Q   L   N   E    E   I   N    L   F   N

901   GGCGCTGTCA  AAACTGTTTC  GTCGAAAGAA  AGTGAAAGCT  GCAATCAAAG  GTTATGTGCG
       Q   A   L   S    K   L   F    R   R   K    K   V   K   A    A   I   K    G   Y   V

961   AAAACTCGAA  ATCACGTACA  ACAAAGAGCG  AGACGACTAT  AATCCGCATT  TTCACTTGAT
       R   K   L   E    I   T   Y    N   K   E    R   D   D   Y    N   P   H    F   H   L

1021  TTTGGCAGTG  AATAAGTCGT  ATTTTACTAA  CCCTAGATAT  TATATCAATC  AAGTAGAATG
       I   L   A   V    N   K   S    Y   F   T    N   P   R   Y    Y   I   N    Q   V   E

1081  GCTGGATCTA  TGGCGTGATG  TGACGGGAAA  AACAGGAGTT  AATCCGGACG  GCACCGACGA
       W   L   D   L    W   R   D    V   T   G    K   T   G   V    N   P   D    G   T   D

1141  GATTACACAG  CTGGATATTC  GCAAGGTTAA  AGGGTTCCAA  CAAGAGAAAG  CAGTCTTGGA
       E   I   T   Q    L   D   I    R   K   V    K   G   F   Q    Q   E   K    A   V   L

1201  AGTCGCAAAA  TATTCGGCCA  AAGACTTTGA  AATGACGGAA  AACCAAGCAG  TATTCGATAC
       E   V   A   K    Y   S   A    K   D   F    E   M   T   E    N   Q   A    V   F   D

1261  GTTTTATTTT  GCAATGAAAG  GTCGCCAGTT  GATTACGTTC  AACGGTGTGT  TCAAGACTAC
       T   F   Y   F    A   M   K    G   R   Q    L   I   T   F    N   G   V    F   K   T

                         rep371
1321  AAAAAGAAAT  TTGAGTCTGG  TGCTTTAGAC  CGATACAAAC  AAAAAGACGA  GAATGATTAT
       T   K   R   N    L   S   L    V   L   -
```

A

```
                -------MSE--NR------------VPTRMHELMMGCGSYLEFIATVDKEKKKLVQAHFC
ER371
C194        MCYNMEKYTEKKQRNQVFQKFIKRHIGENQMDLVEDCNTFLSFVADKTLEKQKLYKANSC
SK89        ----------------------------MDLVKNCNTFLSFVTDKKFEKKKLYKSNPC
                                           .*.   *...* *..    **.**  ...  *
```

```
            KNRFCPLCAWRKARKDAMMLSIMMQAIAQEKQYEFLFMTLTTPNVKGNQLNEEINLFNQA
ER371       KNRFCPVCAWRKARKDALGLSLMMQYIKQQEKKEFIFLTLTTPNVMSDELENEIKRYNNS
C194        KNRFCPVCAWRKARKDALGLSLMMQYIKQEEDKQFIFLTLTTPNVTDEHLESEIKNYNHA
SK89        ******.**********. **.*** * *.  .*.*.*******  *  **  .* .
```

```
            LSKLFRRKKVKAAIKGYVRKLEITYNKERDDYNPHFHLILAVNKSYFTNPRYYINQVEWL
ER371       FRKLIKRKKVGSVIKGYVRKLEITYNKKRDDYNPHFHVLIAVNKSYFTDKRYYISQQEWL
C194        FQKMFKRKKVNAITKGYVRKLEITYNSKRDDYNPHFHVLMAVNKSYFKDTKAYISQKEWL
SK89        *. .**** .  ************ *********...******* . ** * ***
```

```
            DLWRDVTGKTGVNPDGTDEITQLDIRKVKGFQQEKAVLEVAKYSAKDFEMTENQAVFDTF
ER371       DLWRDVTGIS--------EITQVQVQKIRQNNN-KELYEMAKYSGKDSDYLINKSKSL--
C194        NLWRDVTGIS--------EITQVHVQKIKQNSN-KELYEMAKYSGKDSDYLINQKVFDAF
SK89        ******* .         ****. ..*..  . * . *.**** ** .  *.    .*
```

```
            YFAMKGRQLITFNGVFKTTKRNLSLVL---------------------------------
ER371
C194        ---------------------------------------------------------
SK89        YKSLKGKQILVYSGLFKDARNKLKNGDLDYLKDIDPTEYIYQIFYIWNQKEYLASEIYDL
            *  ..**.*...  .  *.**  ..  *
```

```
ER371       --------------------
C194        --------------------
SK89        TEQEKEKINHQMINEIEEEK
```

B

```
ER371  252      TTTTGTTATTTTATTTGAGATTTTTGAGCGCACGGAAAAA-GA-TCTCT-TCTTTTCTTG
C194   1053(C)   TTTTG-T-GTTTTTTTAAAGGATTTGAGCGTAGCGAAAAATCCTTTTCTTTCTTATCTTG
SK89   1415(C)   GATTGATGGTTTTTGAAAGGTTTTTGAAACGAGTTGAAAACGAGTTTCT-TCTTGTCTTG
                 *** * *** * * * ***** * **** * *** **** *****
```
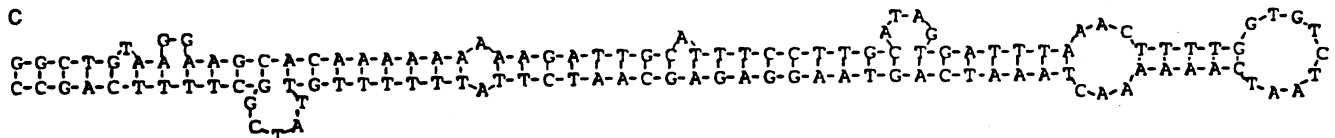
```
ER371   ATAATATAGGC-AACTATTTTCGGCGAAAAAATAGGGGTTTGACC-T---AGGTCAAACC
C194    ATAATA-AGGGTAACTATTGCCGGCGAGGC--TA---GTT--ACCCTTA-AGTT-A----
SK89    ATAATA-AGGGTAACTATTCCCGGCGATGT-ATA--------ACCCTTTCAATTCA----
        ****** **** ******* ****** ** *** * * *
```

```
ER371   CTTGG---GGCT-GTAAGGAAGCACA      385
C194    -TTGGTATGACTGGTTTT-AAGCGCA      926(C)
SK89    -TTGGTATAACTGGTTT-GAAGCACA     1287(C)
        ****  ** ** **** **
```

C

Fig. 2A–C Computer analyses of pER371 replication sequences. The CLUSTAL-W program was used to perform alignment analyses. * A perfect match; • conservative substitution, in the case of amino-acid sequence alignment. Dashes have been introduced to produce optimal matching. The sequences of pC194 (accession no. J01754) and pSK89 (accession no. M37889) were retrieved from GenBank. A Multiple amino-acid sequence alignment of Rep371 with the replication proteins of pC194 and pSK89. B Multiple nucleic-acid sequence alignment of the ORI regions of pER371, pC194 and pSK89. Numbers indicate the positions of the first and last base of the fragments in their respective plasmid sequences as retrieved from the database; (C) the complementary strand of the plasmids. The conserved nick-site sequences of ORI are underlined. C Possible hairpin structure of the M-O region of pER371. The DNASIS for Windows (v. 2.1) program (Hitachi Software Engineering Ltd.) was used to predict secondary-structure formation from the putative M-O of pER371 (nucleotides 369–494). The calculated ΔG value is −228 kJ/mol

quence that is highly homologous to the plus-strand origins (ORI) of pC194 and pSK89 (Fig. 2B). Located between this ORI(pER371) and the promoter region of

rep371 is a palindromic sequence about 140 nucleotides long. Stem-loop analysis of this sequence produced a hairpin structure depicted in Fig. 2C. Even though the sequence of this palindrome shows little homology to the palA or palB elements of the other gram-positive plasmids (Novick 1989), its location and capability of forming hairpin structure suggest that it is likely the minus-strand origin (M-O) of pER371.

Analysis of sequence outside the replication functions

Two additional ORF coding for putative polypeptides with 104 (ORF2) and 86 (ORF3) amino acid residues, were defined by computer-programmed prediction (Fig. 1A). A BLASTP homology search, using the hypothetical polypeptides as queries, however, did not identify any known proteins in the GenBank database.

The absence of recognizable promoter and Shine-Dalgarno consensus sequences upstream from these ORF further suggests that they are not *bona fide* genes with physiological function.

Submission of the nucleotide sequence of pER371, excluding the replicon region, to BLASTN analysis resulted in a list of plasmids almost exclusively of lactobacillus origin. These plasmids contain a previously unreported, highly conserved nucleotide sequence (CNS) that is homologous to the pER371 region flanked by nucleotides 2443–2639. The relative location of this homologous sequence with respect to the replication functions seems to be preserved in this group of plasmids, as it was found 150–300 bp upstream from the ORI region. Multiple sequence alignment of the CNS showed that the *S. thermophilus* sequence contained an additional intervening segment of about 85 bp, not found in *lactobacillus* plasmids pLH2 and pLC2 and *leuconostoc* plasmid pCI411 (Fig. 3A). Three other *lactobacillus* plasmids pLP1, p353-2, and pVKM1311, however, retained remnants of this intervening sequence. BLASTN analysis of the entire intervening segment of pER371 did not reveal any homologous plasmid-related sequences. Furthermore, when sequences of the regions flanking the CNS of pER371 were subjected to BLASTN analyses, the results did not reveal any homology to sequences of plasmid origin, indicating that the sequence similarity to the lactobacillus plasmids is limited to the CNS region. Secondary structure analysis of the pER371 CNS sequence showed that a hairpin structure could possibly form with $\Delta G = -229$ kJ/mol (Fig. 3B). This hairpin formation feature was also observed with the other CNS regions (data not shown).

Discussion

The complete sequence of pER371 is determined and analyzed in this study. This plasmid has been used to construct cloning vectors that are useful for heterologous gene expression in the thermo-tolerant lactic acid bacterium, *S. thermophilus* (Solaiman and Somkuti 1997). The sequence data show that the G + C composition of pER371 is 38.16 mol%. This value is within the range of the reported G + C contents (37%–40%) of the *S. thermophilus* chromosomal DNA (Hardie and Whiley 1992), ensuring that the plasmid is suited for stable maintenance in this host. Interestingly, homology analysis of the replication function of the streptococcal pER371 showed that it is closely related to those of the staphylococcal plasmids pC194 and pSK89. However, the G + C contents of the genes coding for the replication proteins of the two bacterial classes are markedly different. The *rep371* gene has a G + C composition of 40.32 mol%, while the corresponding values for the *rep* genes of pC194 and pSK89 are 26.23 mol% and 27.07 mol% respectively. The reported G + C contents of *S. aureus* chromosomal DNA is in the range of 30–39 mol% (Kloos et al. 1992). Apparently, a similar replication function is being coded by genes, having different nucleotide compositions, that closely match the chromosomal DNA property of the respective hosts.

Among the three putative ORF identified in pER371, only *rep371*, represented by the ORF1, encodes a putative protein involved in plasmid replication. A traditional start codon (ATG) could be found at nucleotide 632, but because this codon is too far from the clearly identifiable Shine-Dalgarno sequence located at nucleotide 591; a rare start codon (TTG) at position 605 was assigned to the Rep371 protein instead. This assignment places the start codon at the conventional 8-bp downstream location from the Shine-Dalgarno sequence. Although TTG is a rare start codon, it has, nevertheless, been found in several coding sequences of Gram-positive bacteria (Dalphin et al. 1996), including the *S. thermophilus epsB* (Stingele et al. 1996) and *cpsE* (Griffin et al. 1996) genes of the exopolysaccharide synthesis pathway and the *pbp2b* gene coding for penicillin-binding protein 2b (Stingele and Mollet 1996).

While the replication function of pER371 is closely related to the pC194-class plasmids, the results of a BLAST homology search and of a protein-sequence-alignment analysis indicate that its Rep371 gene product is phylogenetically distant from the two previously reported replication proteins of *S. thermophilus*, i.e., RepA of pST1 (Janzen et al. 1992) and RepS of pST1-No.29 (Hashiba et al. 1993). This implies that pER8 (2.0 kb) (Somkuti and Steinberg 1986a), having a restriction map similar to that of pST1 (2.1 kb), is also unrelated to pER371. It is expected that pER371 and pER8 should be compatible in the *S. thermophilus* host. Vectors developed from pER371 should thus provide invaluable alternatives to the pER8-based cloning vectors that have proven useful for expressing heterologous genes in *S. thermophilus* (Solaiman and Somkuti 1997; Somkuti et al. 1993, 1995).

Sequence features identified in this study suggest that pER371 belongs to a family of single-stranded DNA plasmids found in many Gram-positive bacteria (Gruss and Ehrlich 1989; Novick 1989). As with other single-stranded DNA plasmids, we observed an occasional instability problem with cloning vectors constructed from pER371 (Solaiman and Somkuti 1997; unpublished data). The *S. thermophilus* plasmid pER371, however, does not contain the recombination sites (RS$_A$ and RS$_B$) that are commonly found in the rolling-circle type staphylococcal plasmids (Novick et al. 1984). This observation indicates that pER371 did not evolve by the recombination of genetic elements similar to those observed with the comparable staphylococcal plasmids (Novick et al. 1984). This study also uncovers a previously unknown CNS element that is highly conserved in a *Leuconostoc lactis* and several lactobacillus plasmids. Sequence comparison of these CNSs shows that they might have derived from a common evolutionary origin that has undergone various degrees of recombination in *S. thermophilus*, *L. lactis*, and the

A

```
pER371   2474  ATAGATAAGG TGCAGTCAAC GGTAAATCGA ACTAGAAGCT CTTAAATGGC
pLP1     1763  CTATCTAACT GTCTGTCAAC GGTAAATCGA CTTAGAGGGG CTTA------
p353-2    389  TTATTTAACT GTCTGTCAAC GGTAAATCGA CGTAGATAGC CTTA------
pVKM1311  695  CTATTTAACC GTCTGTCAAC GGTAAATCGA ATTAAAGGGG CTTG------
pLH2     2565  ATAATTTACT GTCCGTCAAC GGTAAATCGA CGTAGAACGG CTT-------
pCI411    918  ATAATTTACT GTCCGTCAAC GGTAAATCGA CGTAGAACGG CTT-------
pLC2      564  ATAATTTACT GTCCGTCAAC GGTAAATCGA CGTAGAACGG CTT-------

pER371         CCTTAAAATC GATTCTGAGC GTTTTTAGGT TTAATGCTTT AAAAAGCAAG
pLP1           ---------- ----TTGAGC ---------- ---------- ----------
p353-2         ---------- ----TTGAGC ---------- ---------- ----------
pVKM1311       ---------- ----CTG--C ---------- ---------- ----------
pLH2           ---------- ---------- ---------- ---------- ----------
pCI411         ---------- ---------- ---------- ---------- ----------
pLC2           ---------- ---------- ---------- ---------- ----------

pER371         GAAGAATAAT CAAATGGTTT AGAGGGC-AT TTTAGAGCTT CTGAGAGGCT
pLP1           ---------- -------CTT A-CAGGCGAT ATTAGCCCCT CTTGGAGGCT
p353-2         ---------- -------CGT A-CAGGCGAA ATTAGACTAT CTAGGAGGCT
pVKM1311       ---------- -------TTT AGCAG----- -TTAGCCCCT TTTTGAGGCT
pLH2           ---------- ---------- ---------- -TTAGCCGTT CTAGGAGGCT
pCI411         ---------- ---------- ---------- -TTAGCCGTT CTGGGAGGCT
pLC2           ---------- ---------- ---------- -TTAGCG--T CTGGGAGGCT

pER371         TTAAGGAGTT GACGGCT  2639
pLP1           TTAAGGAGTT GATAGAC  1861
p353-2         TTAAGGAGTT GATAGAC   487
pVKM1311       TTAAGGAGTT GACAGAT   786
pLH2           TTAAGGAGTT GACAGAC  2643
pCI411         TTAAGGAGTT GACGGAC   996
pLC2           TTAAGGAGTT GACGGAC   640
```

B
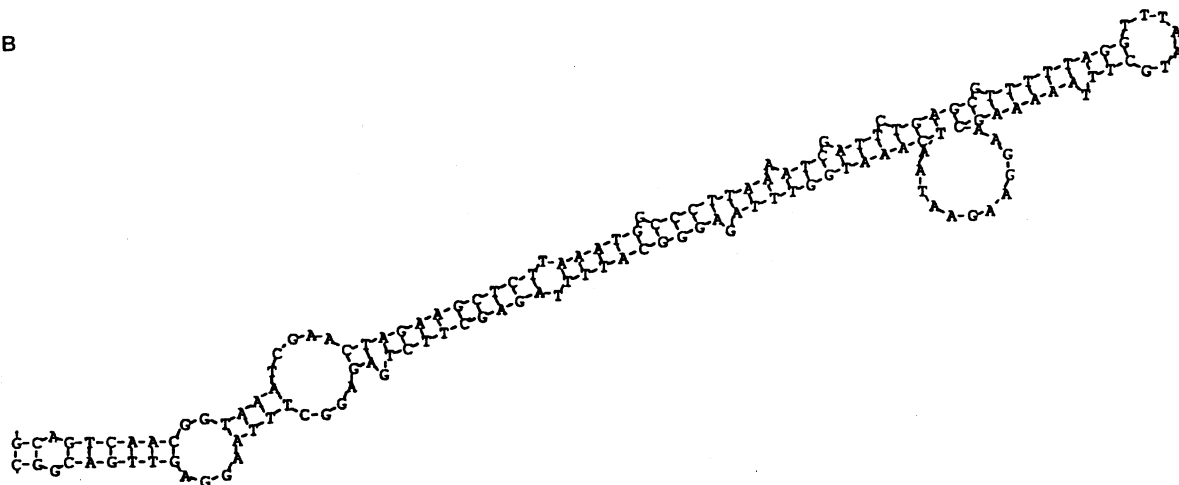
Fig. 3A,B Analysis of the lactococcus-like conserved nucleotide sequence (CNS). **A** Multiple sequence alignment of CNS from *Streptococcus thermophilus* plasmid pER371, *Lactobacillus plantarum* pLP1 (GenBank accession no. M31223), *L. pentosus* p353-2 (X62347), *L. fermentum* pVKM1311 (X74860), *L. helveticus* pLH2 (X81981), *Leuconostoc lactis* pCI411 (L25529), and *L. curvatus* pLC2 (Z14234). Shaded residues represent at least 50% match among the aligned nucleotides. The numbers indicate the positions of the flanking nucleotides of the segments in their respective plasmids as retrieved from the database. **B** Predicted hairpin structure formed by the sequence (nucleotides 2485–2638) present in the CNS of pER371. The calculated $\Delta G = -229$ kJ/mol

lactobacilli. In addition to the high degree of sequence conservation, the CNS are also characterized by their potential to form hairpin structures. These properties imply that CNS might serve some still unknown genetic regulatory functions, though a functional shuttle vector (pMEU14′-1) has been constructed without this element (Solaiman and Somkuti 1997).

The determination of the complete pER371 sequence has allowed its detailed analysis, leading to the characterization of its replication functions and the discovery of a putative genetic regulatory element, CNS. This

information should be invaluable for the design and construction of the second-generation pER371-based food-grade vectors for the genetic transformation of *S. thermophilus* and other dairy-food fermentation microorganisms.

References

Altschul SF, Gish W, Miller W, Myers EG, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struchl K (1987) *E. coli*, plasmids, and bacteriophages. In: Current protocols in molecular biology. Wiley-Interscience, New York, pp 1.7.1–1.7.4

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7: 1513–1522

Constable A, Mollet B (1994) Isolation and characterization of promoter regions from *Streptococcus thermophilus*. FEMS Microbiol Lett 122: 85–90

Dalphin ME, Brown CM, Stockwell PA, Tate WP (1996) TRANSTERM – a database of translational signals. Nucleic Acids Res 24: 216–218

Griffin AM, Morris VJ, Gasson MJ (1996) The *cpsABCDE* genes involved in polysaccharide production in *Streptococcus salivarius* ssp. *thermophilus* strain NCBF 2393. Gene 183: 23–27

Gruss A, Ehrlich SD (1989) The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. Microbiol Rev 53: 231–241

Hardie JM, Whiley RA (1992) The genus *Streptococcus* - oral. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer K-H (eds) The prokaryotes, vol 2, 2nd edn. Springer, New York Berlin Heidelberg, pp 1421–1449

Hashiba H, Takiguchi R, Joho K, Aoyama K, Hirota T (1993) Identification of the replication region of *Streptococcus thermophilus* No. 29 plasmid pST1. Biosci Biotechnol Biochem 57: 1646–1649

Herman RE, McKay LL (1985) Isolation and partial characterization of plasmid DNA from *Streptococcus thermophilus*. Appl Environ Microbiol 50: 1103–1106

Horinouchi S, Weisblum B (1982) Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. J Bacteriol 150: 815–825

Jacobs MF, Tynkkynen S, Sibakov M (1995) Highly bioluminescent *Streptococcus thermophilus* strain for the detection of dairy-relevant antibiotics in milk. Appl Microbiol Biotechnol 44: 405–412

Janzen T, Kleinschmidt J, Neve H, Geis A (1992) Sequencing and characterization of pST1, a cryptic plasmid from *Streptococcus thermophilus*. FEMS Microbiol Lett 95: 175–180

Kiewiet R, Kok J, Seegers JFML, Venema G, Bron S (1993) The mode of replication is a major factor in segregational plasmid instability in *Lactococcus lactis*. Appl Environ Microbiol 59: 358–364

Kloos WE, Schleifer K-H, Götz F (1992) The genus *Staphylococcus*. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer K-H (eds) The prokaryotes, vol 2, 2nd edn. Springer, New York Berlin Heidelberg, pp 1369–1420

Krabbe M, Zabielski J, Bernander R, Nordstrom K (1997) Inactivation of the replication- termination system affects the replication mode and causes unstable maintenance of plasmid R1. Mol Microbiol 24: 723–735

Littlejohn TG, DiBerardino D, Messerotti LJ, Spiers SJ, Skurray RA (1991) Structure and evolution of a family of genes encoding antiseptic and disinfectant resistance in *Staphylococcus aureus*. Gene 101: 59–66

Mercenier A (1990) Molecular genetics of *Streptococcus thermophilus*. FEMS Microbiol Rev 87: 61–77

Moineau S, Walker SA, Holler BJ, Vedamuthu ER, Vandenbergh PA (1995) Expression of *Lactococcus lactis* phage resistance mechanism by *Streptococcus thermophilus*. Appl Environ Microbiol 61: 2461–2466

Mollet B, Knol J, Poolman B, Marciset O, Delley M (1993) Directed genomic integration, gene replacement, and integrative gene expression in *Streptococcus thermophilus*. J Bacteriol 175: 4315–4325

Novick RP (1989) Staphylococcal plasmids and their replication. Annu Rev Microbiol 43: 537–565

Novick RP, Projan SJ, Rosenblum W, Edelman I (1984) Staphylococcal plasmid cointegrates are formed by host- and phage-mediated general *rec* systems that act on short regions of homology. Mol Gen Genet 195: 374–377

Renault P, Corthies G, Goupil N, Delorme C, Ehrlich SD (1996) Plasmid vectors for Gram-positive bacteria switching from high to low copy number. Gene 183: 175–182

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Seegers JFML, Bron S, Franke CM, Venema G, Kiewiet R (1994) The majority of lactococcal plasmids carry a highly related replicon. Microbiology 140: 1291–1300

Solaiman DKY, Somkuti GA (1993) Shuttle vectors developed from *Streptococcus thermophilus* native plasmid. Plasmid 30: 67–78

Solaiman DKY, Somkuti GA (1997) Characterization of pER371-based *Streptococcus thermophilus-Escherichia coli* shuttle vectors. Biotechnol Lett 19: 595–598

Somkuti GA, Steinberg DH (1986a) Distribution and analysis of plasmids in *Streptococcus thermophilus*. J Ind Microbiol 1: 157–163

Somkuti GA, Steinberg DH (1986b) General method for plasmid DNA isolation from thermophilic lactic acid bacteria. J Biotechnol 3: 323–332

Somkuti GA, Solaiman DKY, Johnson TL, Steinberg DH (1991) Transfer and expression of a *Streptomyces* cholesterol oxidase gene in *Streptococcus thermophilus*. Biotechnol Appl Biochem 13: 238–245

Somkuti GA, Solaiman DKY, Steinberg DH (1993) Cloning of a tyrosinase gene in *Streptococcus thermophilus*. Biotechnol Lett 15: 773–778

Somkuti GA, Solaiman DKY, Steinberg DH (1995) Native promoter-plasmid vector system for heterologous cholesterol oxidase synthesis in *Streptococcus thermophilus*. Plasmid 33: 7–14

Stingele F, Mollet B (1996) Disruption of the gene encoding penicillin-binding protein 2b (pbp2b) causes altered cell morphology and cease in exopolysaccharide production in *Streptococcus thermophilus* Sfi6. Mol Microbiol 22: 357–366

Stingele F, Neeser JR, Mollet B (1996) Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. J Bacteriol 178: 1680–1690

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Vaughan EE, David S, DeVos WM (1996) The lactose transporter in *Leuconostoc lactis* is a new member of the *lacS* subfamily of galactoside-pentose-hexuronide translocators. Appl Environ Microbiol 62: 1574–1582